



مبانی بازیابی اطلاعات و جستجوی وب

نیم سال دوم 1403-1404

تمرین دوم

عنوان: طراحی سیستم تحلیل و بررسی رتبه کلمات کلیدی سایت‌ها

مهلت تحویل:

ارسال با تاخیر 5 روزه امکان‌پذیر است، اما توصیه نمی‌شود.
ارسال دیرهنگام، روزانه 5 درصد جریمه دارد. ارسال بعد از 5 روز تاخیر شامل هیچ نمره‌ای نمی‌شود.

مقدمه

این پروژه یک سیستم کامل برای تحلیل رتبه سایت‌ها در جستجوی گوگل است که با تعریف سایت‌ها و کلمات کلیدی مرتبط شروع می‌شود و با اجرای ربات سلنیوم برای جستجو و جمع‌آوری داده‌ها تکمیل می‌شود. داده‌های جمع‌آوری شده شامل کمترین، بیشترین، و میانگین رتبه کلمات کلیدی و گزارش‌های جامع بصری در داشبورد است. فازهای پروژه به ترتیب عبارتند از تعریف اطلاعات اولیه، تولید ربات جستجو در گوگل و تولید داشبورد نمایش اطلاعات.

بخش اول: تعریف اطلاعات اولیه

هدف: در این بخش می‌خواهیم یک صفحه برای وارد کردن تعدادی کلید واژه و وبسایت‌های نظیر به صورت CRUD طراحی کنیم.

در این بخش، کاربر می‌تواند سایت‌ها و کلمات کلیدی مرتبط با آن‌ها را مدیریت کند. کاربر از طریق یک رابط کاربری ساده می‌تواند سایت‌های جدید را با مشخص کردن نام و دامنه اضافه کرده و سپس کلمات کلیدی مرتبط با هر سایت را وارد کند. همچنین، امکان آپلود فایل اکسل برای وارد کردن تعداد زیادی کلمات کلیدی به صورت دسته‌ای فراهم است. این بخش، پایه اصلی ورود داده‌های پروژه را تشکیل می‌دهد و اطلاعات از طریق فرم‌ها یا فایل اکسل به پایگاه داده ارسال می‌شود تا در مراحل بعدی استفاده شود. جزئیات این بخش در ادامه بیان شده است.

تعریف سایت‌ها:

- هر سایت شامل:
 - نام سایت (مانند: فروشگاه آنلاین کتاب)
 - دامنه سایت (مانند example.com)
- قابلیت افزودن، حذف، و ویرایش سایت‌ها از طریق رابط کاربری و REST API فراهم شود.

تعریف کلمات کلیدی:

- امکان تعریف کلمات کلیدی با قابلیت اختصاص چند سایت مرتبط (یک کلمه کلیدی ممکن است برای چند سایت بررسی شود).
- امکان ورود این داده‌ها از فایل اکسل (قالب پیشنهادی):

keyword	sites
خرید کتاب	bookshop.com, bookstore.com, idebook.ir
گوشی هوشمند	mobilestore.com

تصویر 1- نمونه فایل اکسل ورودی

فرم ورود کلمات کلیدی

سایت مرتبط:

انتخاب سایت

کلمه کلیدی:

کلمه کلیدی را وارد کنید

آپلود فایل اکسل (اختیاری):

فایلی انتخاب نشده است

انتخاب فایل

ارسال

تصویر 2- فرم نمونه ورودی

توجه: انتخاب کلید واژه ها و وب سایت ها به عهده تیم ها است. دقت کنید که حداقل 10 کلید واژه و برای هر کلید واژه حداقل 2 وبسایت وجود داشته باشد.

بخش دوم: تولید ربات جست و جو در گوگل

هدف: انجام جست و جو برای کلید واژه و وب سایت مورد نظر

این بخش شامل توسعه یک ربات وب با استفاده از ابزاری مانند Selenium است که وظیفه جستجوی کلمات کلیدی در گوگل و یافتن رتبه سایت‌های مرتبط را بر عهده دارد. ربات از طریق REST API کلمات کلیدی و سایت‌های مرتبط را دریافت می‌کند، آن‌ها را در گوگل جستجو کرده و رتبه سایت‌ها را تا عمق مشخصی (مانند صفحه ۱۰) استخراج می‌کند. همچنین، قابلیت کلیک تصادفی روی نتایج مرتبط و ماندن در صفحات برای شبیه‌سازی رفتار کاربر وجود دارد که به بهبود نتایج سئو کمک می‌کند. داده‌های به‌دست‌آمده به صورت مرتب به پایگاه داده ارسال می‌شوند.

وظایف اصلی ربات

1. دریافت کلمات کلیدی و سایت‌های مرتبط از طریق REST API
2. جستجوی کلمات کلیدی در گوگل تا رتبه حداکثر تعیین‌شده بصورت کانفیگ (مثلاً 50)
3. جمع‌آوری رتبه سایت‌های مرتبط برای هر کلمه کلیدی
4. ذخیره اطلاعات جمع‌آوری‌شده شامل کمترین، بیشترین، و میانگین رتبه هر کلمه کلیدی
5. مشاهده لاگ ربات‌ها (آی پی دسترسی، برچسب زمانی و نوع درخواست REST API)
6. کلیک تصادفی روی سایت مرتبط (در صورت یافتن)، ماندن روی صفحه برای مدت زمان مشخص، و ادامه جستجو (اختیاری و نمره اضافه)
7. جمع‌آوری کلمات کلیدی پیشنهادی گوگل به ازای هر کلمه کلیدی و ذخیره در یک جدول جداگانه (اختیاری و نمره اضافه)

پیکربندی

- زمان‌بندی اجرای ربات (مانند هر روز یا چند بار در روز)
- ارسال کلمات با REST API بصورت تصادفی بوده و برای هر کلمه کلیدی یک آستانه روزانه تعریف می‌شود که در صورت رسیدن به آستانه دیگر آن کلمه برای سرچ داده نمی‌شود.
- تنظیم حداکثر رتبه بررسی (مثلاً 50)
- مدت زمان ماندن روی سایت پس از کلیک تصادفی

مثال داده‌های جمع‌آوری شده توسط ربات:

تعداد بررسی	میانگین رتبه	بیشترین رتبه	کمترین رتبه	دامنه سایت	کلمه کلیدی
4	7.5	12	3	bookshop.com	خرید کتاب
4	10.3	15	5	bookstore.com	خرید کتاب
3	4.5	8	1	mobilestore.com	گوشی هوشمند

توجه: ربات هیچگونه دسترسی ای برای تغییر ندارد و در صورتی که کلمه ای برای آن روز باقی نمانده باشد پیغام مناسب سمت ربات ارسال شده که به معنی پایان کلمات و عدم درخواست مجدد در آن روز است.

بخش سوم: تولید داشبورد نمایش اطلاعات

هدف: نمایش داشبورد اطلاعاتی

این بخش رابط کاربری بصری برای مشاهده و تجزیه و تحلیل داده‌های ثبت شده است. داشبورد شامل نمودارهایی برای نمایش روند تغییر رتبه کلمات کلیدی در روزهای مختلف، تعداد دفعات بررسی، و مقایسه رتبه سایت‌ها بر اساس کلمات کلیدی است. همچنین، یک جدول اطلاعات دقیق هر کلمه کلیدی شامل کمترین، بیشترین، و میانگین رتبه به همراه تعداد بررسی‌ها را نمایش می‌دهد. این داشبورد به کاربران امکان می‌دهد تا عملکرد سایت‌های خود را تحلیل کرده و استراتژی‌های سئو را بر اساس داده‌های دقیق تنظیم کنند.

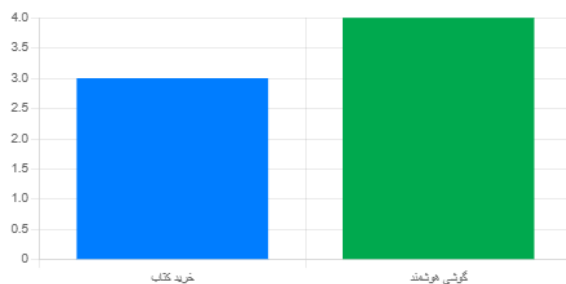
ویژگی‌های داشبورد:

- **نمایش جدول داده‌ها:**
 - کمترین، بیشترین، میانگین رتبه کلمات کلیدی به تفکیک سایت‌ها.
- **نمودارها:**
 - **نمودار خطی:** نمایش تغییرات کمترین، بیشترین و میانگین رتبه کلمات کلیدی در بازه زمانی مشخص.
 - **نمودار میله‌ای:** تعداد بررسی‌های انجام شده برای هر کلمه کلیدی.
- **فیلترگذاری:**
 - فیلتر بر اساس بازه زمانی، سایت، یا کلمه کلیدی.

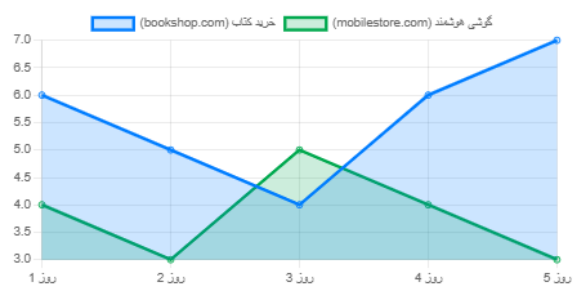
توجه: برای هر کلید واژه و وب سایت چندبار و در ساعت‌ها و روزهای متفاوت جست و جو را انجام دهید تا بهترین رتبه و بدترین رتبه متمایز باشند و علاوه بر این بتوانید نمودار چند روزه رتبه بندی را رسم کنید.

داشبورد تحلیلی

تعداد بررسی‌های انجام شده



تغییرات رتبه کلمات کلیدی



جدول داده‌های کلمات کلیدی

کلمه کلیدی	سایت مرتبط	کمترین رتبه	بیشترین رتبه	میانگین رتبه	تعداد بررسی
خرید کتاب	bookshop.com	2	10	6.3	3
گوشی هوشمند	mobilestore.com	1	8	4.5	4

تصویر 3- نمونه ساده ای از داشبورد اطلاعاتی

شیوه تحویل

تمام عملیات پروژه، از مدیریت داده‌ها تا دریافت نتایج از ربات و نمایش آن‌ها در داشبورد، از طریق REST API انجام می‌شود. رابط کاربری در بخش‌های مختلف به صورت ماژولار طراحی شده و از API برای ارسال، دریافت، و بروزرسانی داده‌ها استفاده می‌کند. طراحی رابط کاربری به صورت مینیمال و کاربرپسند انجام شده است، اما با افزودن جزئیات بصری بیشتر و استفاده از فریمورک طراحی رابط کاربری، می‌توانید نمره اضافه دریافت کنید. این ساختار باعث جداسازی کامل بخش‌های Front و Back می‌شود و انعطاف‌پذیری در توسعه و نگهداری پروژه را تضمین می‌کند.

ساختار پروژه:

• Frontend:

- طراحی رابط کاربری ساده با امکان:
 - افزودن سایت‌ها و کلمات کلیدی
 - مشاهده اطلاعات داشبورد
 - نمایش وضعیت اجرای ربات

• Backend:

- پیاده‌سازی REST API با استفاده از فریمورک‌هایی مانند Flask یا Django
- مدیریت پایگاه داده و اجرای ربات سلنیوم
- مستندات API برای عملیات CRUD (ایجاد، خواندن، به‌روزرسانی و حذف) و دریافت نتایج.

• ربات جستجو:

- طراحی ربات با ابزاری مثل سلنیوم

امکانات اضافی (برای دریافت نمره بیشتر):

1. طراحی رابط کاربری حرفه‌ای:

- استفاده از فریمورک‌های مدرن مانند React.js یا Vue.js برای طراحی رابط کاربری حرفه‌ای
- امکانات حرفه‌ای برای نمایش داشبورد

2. قابلیت‌های اضافی ربات:

- پس از یافتن سایت مرتبط در گوگل، ربات به صورت تصادفی روی لینک کلیک کند و برای مدت زمان مشخصی روی سایت بماند.
- جمع آوری کلمات کلیدی پیشنهادی گوگل به ازای هر کلمه کلیدی و ذخیره در یک جدول جداگانه

3. گزارش‌گیری خودکار و اعلان:

- نمایش میزان صعود یا نزول کلماتی که بیشتر یا کمتر از یک آستانه بوده است.
- نمایش کلماتی که فاصله زیادی از کلمات مورد جستجوی سایر افراد است.
- تولید گزارش‌های تحلیلی
- ارسال گزارش به ایمیل کاربران

موارد تحویلی:

- هر قسمت پروژه (فرانت‌اند، بک‌اند و ربات) بصورت جداگانه و با یک readme بایستی باشد.
- امکان اجرای یکپارچه همه اجزا در زمان تحویل پروژه فراهم باشد.
- اطلاعات چند روز در داخل دیتابیس موجود و امکان نمایش داشبورد فراهم باشد.
- یک داکيومنت از خروجی‌ها به همراه عملیات انجام شده بنویسید.

نمره‌دهی پروژه:

- پیاده‌سازی دقیق REST API ها: 40%
- پیاده‌سازی دقیق ربات: 40%
- رابط کاربری ساده و کاربردی: 20%
- طراحی رابط کاربری حرفه‌ای: 20%+ نمره اضافی.
- قابلیت اضافی ربات: 10%+ نمره اضافی.
- گزارش‌گیری و اعلان‌ها: 10%+ نمره اضافی.